AD-A078 283 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 12/1

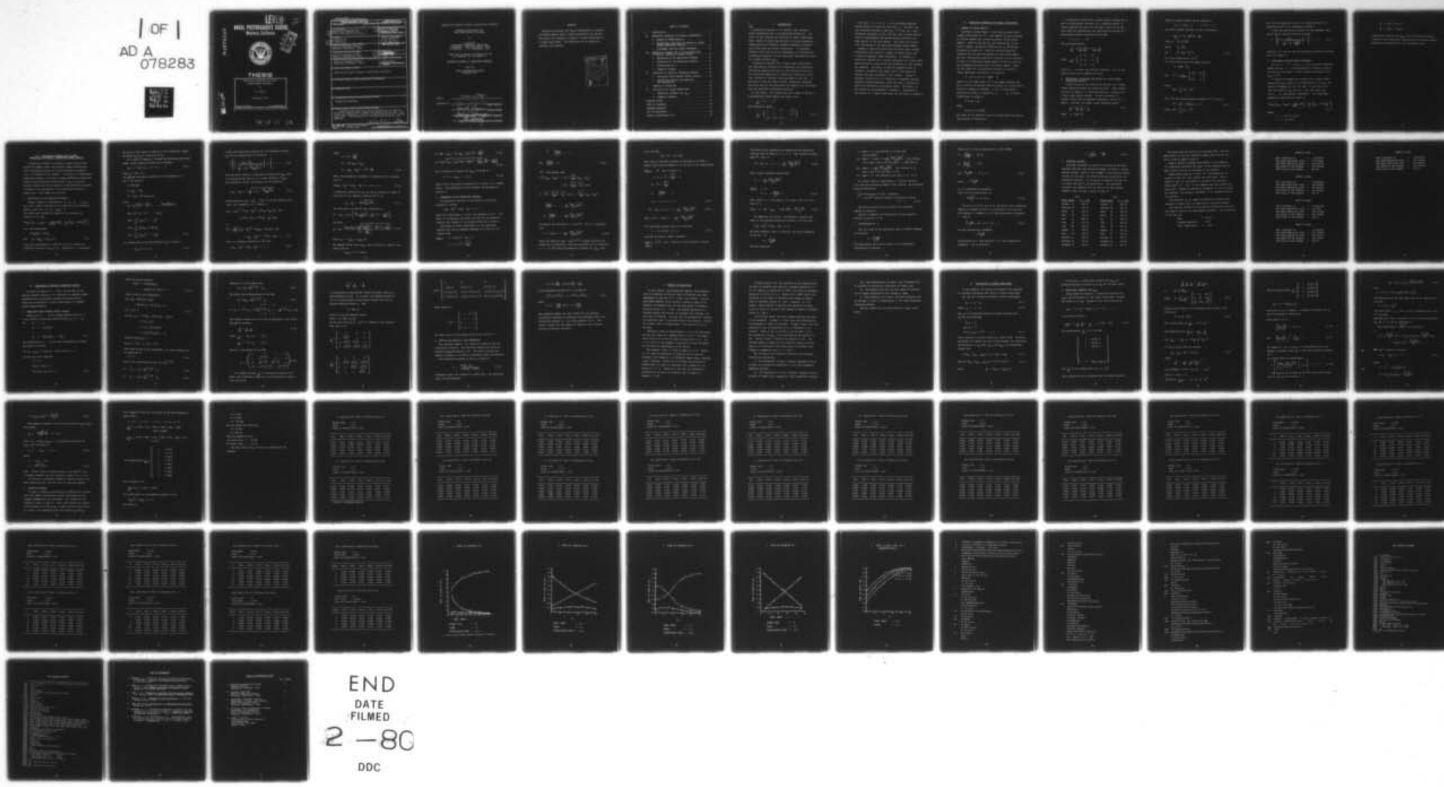
PREDICTION INTERVALS FOR FIRST ORDER MARKOV PROCESSES.(U)

SEP 79 T S MURTHY

UNCLASSIFIED NL

1 OF 1

AD A 078283



END

DATE FILMED

2-80

DDC

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

PREDICTION INTERVALS FOR FIRST
ORDER MARKOV PROCESSES

by

T S Murthy

September 1979

Thesis Advisor: T Jayachandran

Approved for public release; distribution unlimited

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 Prediction Intervals for First Order Markov Processes.		5. ~~TYPE OF REPORT & PERIOD~~ COVERED 9 Master's Thesis, ~~September 1979~~
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 T. S. Murthy		8. CONTRACT OR GRANT NUMBER(s) 12 71
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE 11 Sep[redacted]79
		13. NUMBER OF PAGES 70
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Prediction Intervals

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Prediction Intervals for future observations in serially correlated samples from a normal distribution are derived. The results are extended to predict a future observation in a linear trend model. The properties of the prediction intervals are examined.

251 450 Gm

Prediction Intervals for
First Order Markov Processes

by

T S Murthy
Lieutenant Commander, Indian Navy
B.Sc(Hons)., Utkal University, 1963
B.Tech(Hons)., I I T Kharagpur, 1966

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1979

Author ______________________________

Approved by: ______________________________
Thesis Advisor

____
Second Reader

____
Chairman, Department of Operations Research

____
Dean of Information and Policy Sciences

ABSTRACT

Prediction Intervals for future observations in serially correlated samples from a normal distribution are derived. The results are extended to predict a future observation in a linear trend model. The properties of the prediction intervals are examined.

TABLE OF CONTENTS

I. INTRODUCTION

A prediction interval is an interval that contains a future observation with a pre-specified probability. The limits of the interval are functions of known observations from a family of known distributions with given parameters. Though prediction intervals resemble confidence intervals, they differ from the latter conceptually. A confidence interval covers the value of a parameter of a distribution. A prediction interval, on the other hand, encloses the value of a random variable.

A prediction interval for a single future observation and simultaneous intervals for independent and identically distributed samples from $N(\mu,\sigma^2)$ were derived by Hahn [2]. For correlated samples with a prescribed correlation structure, prediction intervals were developed by Choi [3]. It was shown by him that many of Hahn's prediction intervals are valid even for the case where the samples were correlated with the specified correlation structure.

In this thesis, the observations are assumed to belong to a multivariate normal family with mean vector

$$\underset{n\times 1}{\underline{\mu}} = (\mu,\mu,\ldots,\mu)^T$$

and correlation matrix

$$\underset{n\times n}{\underline{V}} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 \end{pmatrix} \qquad (1\cdot 1)$$

Let $X_t (t=\ldots-2,-1,0,1,2\ldots)$ be a stationary Gaussian markov process of order one with $E(X_t) = \mu$ for all t and the covariance between X_t and $X_{t+k} = \sigma^2\rho^k$ for all t and k. Stochastic processes of this type are used to model certain types of oceanographic and metereological phenomena and some aspects of the stock market. Any finite sub process will then belong to the above multivariate normal family.

Chapter 2 deals with prediction intervals from a normal distribution and contains a summary of known results and a review of general linear hypothesis. In chapter 3 a prediction interval for a future observation in a first order markov process is derived and the properties of this interval are examined. The effect of various parameters such as variance, correlation and sample size on the prediction interval are also discussed in this chapter. Chapter 4 contains a discussion of a simulation to generate serially correlated random variables, which are used to empirically verify the theoretical conclusions drawn. The results of the simulation are presented in chapter 5. Extensions of the results to a linear trend model are discussed in chapter 6.

II. PREDICTION INTERVALS FOR NORMAL DISTRIBUTION

A. SUMMARY OF KNOWN RESULTS

Consider a random sample of size $N=n_i+n_f$ drawn from a normal distribution $N(\mu,\sigma^2)$. The samples n_i form the group of initial samples and n_f form the group, called future samples. Based on the mean $\overline{X}_{n_i}$ and variance $S^2_{n_i}$ of the initial samples, prediction intervals for $\overline{X}_n$ and S^2_n were obtained by Hickman [6]. Hahn [2] derived prediction intervals for the mean and variance of the second sample and also simultaneous prediction intervals for the variance of each of k additional random samples of size n_f based on the information obtained from the initial sample. Hahn's [2] two sided 100 r % prediction interval to contain a single additional observation X is given by

$$\overline{Y} \pm t[n-1;(1+r)/2]\ (1+\tfrac{1}{n})^{\frac{1}{2}} \cdot S$$

where $\overline{Y}$ is the sample mean, S^2 is the sample variance and $t[n-1,(1+r)/2]$ is obtained from the tables of t distribution with (n-1) degrees of freedom. A 100 r % simultaneous prediction interval to contain $X_1, X_2, \ldots, X_k$, k future observations is given by

$$\overline{Y} \pm R(k,n,r)S$$

where

$$R(k,n,r) = U(1+\tfrac{1}{n})^{\frac{1}{2}}$$

The value of the function U can be obtained from the special tables given in reference 2.

A prediction interval for a single future observation as well as simultaneous intervals for a specified number of future observations have been derived by Choi [3] for the case where the observations are correlated and belong to a multivariate normal distribution with mean

$$\underline{\mu} = (\mu, \mu, \ldots, \mu)^T,$$

and covariance matrix,

$$\underset{n\times n}{\underline{V}} = \tfrac{1}{2}(\underset{n\times n}{\underline{H}} + \underset{n\times n}{\underline{H}^T}) + \alpha(\underset{n\times n}{\underline{I}} - \underset{n\times n}{\underline{E}})$$

where $$\underset{n\times n}{\underline{H}} = \begin{pmatrix} h_1 & h_1 & \cdots & h_1 \\ h_2 & h_2 & \cdots & h_2 \\ h_n & h_n & \cdots & h_n \end{pmatrix}$$

$h_i (i=1,2,\ldots,n)$ and α are positive constants. $\underline{E}$ is an nxn matrix all of whose elements are unity.

B. DEFINITIONS, NOTATIONS AND REVIEW OF SIMPLE LINEAR REGRESSION ANALYSIS

A brief review of the results and notation in a simple linear regression model are presented below. These results are used in chapter 3 in deriving prediction intervals for a first order markov process. By way of notation, a capital letter with an underbar represents a vector or a matrix. Consider the simple linear regression model

$$\underset{n\times 1}{\underline{Y}} = \underset{n\times 2}{\underline{X}}\ \underset{2\times 1}{\underline{B}} + \underset{n\times 1}{\underline{\varepsilon}} \qquad (2\cdot 1)$$

where $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 \underline{I})$

which in scalar notation can be written as

$$Y_i = a + bX_i + \varepsilon_i \qquad i = 1,2,\ldots,n.$$

The least squares estimator for $\underline{B}$ is obtained by

$$\underset{\underline{B}}{\text{Min}}\ Q = (\underline{Y} - \underline{XB})^T(\underline{Y} - \underline{XB}),$$

that is $\quad \hat{\underline{B}} = \underline{S}^{-1}\underline{X}^T\underline{Y}$,

where $\quad \underline{S} = \underline{X}^T\underline{X}$,

and $\quad \hat{\underline{B}} \sim N(\underline{B}, \sigma^2\underline{S}^{-1})$. $\qquad (2\cdot2)$

$\underline{B}$ is the column vector $(a,b)^T$.

It is well known that the random variable

$$\underline{Y} \sim N(\underline{XB}, \sigma^2\underline{I})$$

and $\quad S^{-1} = \dfrac{1}{n\cdot SXX}\begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$,

where

$$SXX = \sum_{i=1}^{n} (X_i - \overline{X})^2 .$$

Let $\hat{\sigma}^2$ be the usual unbiased estimator of σ^2, given by

$$\hat{\sigma}^2 = (SYY - \hat{b}^2 SXX)/(n - 2) , \qquad (2\cdot3)$$

where $\quad SYY = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$.

Let $\hat{Y}$ be the predicted value of the random variable at a specified value of the independent variable x.

A 100(1-α)% prediction interval for the dependent variable Y can be obtained from the relation

$$P\left[\left|\frac{Y-\hat{Y}}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{SXX}}}\right| \leq t\right] = 1 - \alpha \qquad (2\cdot 4)$$

Equation (2·4) will be used in deriving a prediction interval for x in chapter 3.

C. STATIONARY GAUSSIAN MARKOV PROCESSES

Let X_t (t = ...-2,-1,0,1,2,...) be a stationary gaussian markov process, i.e., an autoregressive stochastic process, of order one with $E(X_t) = m$ and covariance between X_t and $X_{t+k} = \sigma^2\rho^k$.

To derive the prediction interval for a single future observation X_{2k+1} based on the observations up to X_{2k}, the results of Ogawara [1] will be used. He has shown that when the values of X_{2k-1}(k=1,2,..,n+1) are fixed, X_{2k} (k=1,2,...,n) are mutually independent. The conditional probability densities are given by

$$f(X_{2k} \mid X_{2k-1}, X_{2k+1}) = \frac{1}{\sqrt{2\pi\sigma_o}} \text{EXP}\left[-\frac{1}{2\sigma_o^2}\left\{X_{2k} - (a+bX_k')\right\}^2\right], \qquad (2\cdot 5)$$

where

$$a = m(1-\rho^2)/(1+\rho^2)$$

$$b = 2\rho/(1+\rho^2)$$

$$\sigma_o^2 = \sigma^2(1-\rho^2)/(1+\rho^2)$$

$$X_k' = (X_{2k-1} + X_{2k+1})/\ 2$$

A prediction interval for X_{2k+1} will be derived in chapter 3 based on this conditional distribution of the even numbered observations conditioned on the odd numbered ones.

III. PREDICTION INTERVAL FOR A FUTURE OBSERVATION IN SERIALLY CORRELATED NORMAL SAMPLES

A prediction interval to contain a single future observation for samples from a multivariate normal distribution with a correlation matrix of the form defined in equation (1·1), is derived in this chapter. In section A a conditional prediction interval to contain a single additional observation based on the conditional distribution of X_{2k} is obtained. Section B deals with the properties of the prediction interval. In section C the dependence of the prediction interval on sample size, ρ and sigma are discussed.

A. DERIVATION OF THE PREDICTION INTERVAL

Assume a sample of observations, $X_1, X_2, X_3, \ldots, X_{2k-1}, X_{2k}$ from a normal distribution with mean zero and covariance matrix $\underline{V}$ such that $\mathrm{Cov}\,(X_t, X_{t+k}) = \rho^k$.
The conditional probability densities, as indicated in equation (2·5), are

$$f(X_{2k} \mid X_{2k-1}, X_{2k+1}) = \frac{1}{\sqrt{2\pi\sigma_o}} \,\mathrm{EXP}\left[-\frac{1}{2\sigma_o^2} \left\{ X_{2k} - (a+bX'_k) \right\}^2 \right]$$

with conditional mean

$$E(X_{2k} \mid X'_k) = a+bX'_k \;,$$

and $$X'_k = (X_{2k-1} + X_{2k-1})/2 \quad . \qquad (3\cdot 1)$$

$E(X_{2k} \mid X'_k)$ considered as a function of X'_k is called the regression function of X_{2k} on X'_k. Graphically, it represents

the locus of the center of gravity of the conditional random variables $X_{2k} \mid X'_k$ as a function of X'_k.

As a result of Ogawara's Theorem the following conditional simple linear regression model may be assumed.

$$X_{2k} = a + bX'_k + \varepsilon_k \ , \ k = 1,2, \ldots, n \tag{3·2}$$

where $\varepsilon_k = N(0, \sigma^2)$.

The maximum likelihood estimators of the parameters a, b and σ^2 are given by

$$\hat{b} = SXY/SXX$$

$$\hat{a} = \bar{x}_{2k} - \hat{b}\,\bar{x}'_k$$

$$\hat{\sigma}^2 = (SYY - \hat{b}^2\, SXX)/(n-2) \ .$$

where

$$\bar{x}'_k = \frac{1}{n}\left(\frac{x_1+x_3}{2} + \frac{x_3+x_5}{2} + \ldots. + \frac{x_{2k-1}+x_{2k+1}}{2}\right) ,$$

$$\bar{x}_{2k} = \frac{1}{n}\left(x_2 + x_4 + \ldots + x_{2k}\right) ,$$

$$SXX = \sum_{k=1}^{n}\left(X_{2k}\right)^2 - n\ \bar{x}'^2_k ,$$

$$SXY = \sum_{k=1}^{n} X'_k\, X_{2k} - n\ \bar{x}'_k\, \bar{x}_{2k} ,$$

$$SYY = \sum_{k=1}^{n}\left(X_{2k}\right)^2 - n\ \bar{x}_{2k}^2 \ . \tag{3·3}$$

At a known value of x'_k the predictor $\hat{x}_{2k}$ is given by

$$\hat{x}_{2k} = \hat{a} + \hat{b}\ \ x'_k \tag{3·4}$$

A 100(1-α)% prediction interval for the dependent variable x_{2k} can be obtained from the relation

$$P\left[\left|\frac{x_{2k} - \hat{x}_{2k}}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x'_k - \bar{x}'_k)^2}{SXX}}}\right| \leq t\right] = 1 - \alpha\ . \qquad (3\cdot 5)$$

From the above relation a prediction interval for x_{2k+1} can be obtained noting that x'_k is a linear function of x_{2k+1}. The inequality inside the brackets can be written as

$$(x_{2k} - \hat{x}_{2k}) \leq t\ \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x'_k - \bar{x}'_k)^2}{SXX}}\ . \qquad (3\cdot 6)$$

Substituting for $\hat{x}_{2k} = (\bar{x}_{2k} - \hat{b}\ \bar{x}'_k) + \hat{b}\ x'_k$ and squaring both sides, the inequality (3·6) reduces to

$$(x_{2k} - \bar{x}_{2k})^2 + \hat{b}^2(x'_k - \bar{x}'_k)^2 - 2\hat{b}(x_{2k} - \bar{x}_{2k})(x'_k - \bar{x}'_k)$$

$$\leq t^2\hat{\sigma}^2(1 + \frac{1}{n}) + t^2\hat{\sigma}^2(x'_k - \bar{x}'_k)^2/SXX$$

or

$$\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)\left(x'_k - \bar{x}'_k\right)^2 - 2\hat{b}(x_{2k} - \bar{x}_{2k})(x'_k - \bar{x}'_k)$$

$$+ (x_{2k} - \bar{x}_{2k})^2 - t^2\hat{\sigma}^2(1 + \frac{1}{n}) \leq 0 \qquad (3\cdot 7)$$

This is a quadratic equation of the form

$$A(x'_k - \bar{x}'_k)^2 + B(x'_k - \bar{x}'_k) + C \leq 0\ . \qquad (3\cdot 8)$$

where

$$A = \hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}$$

$$B = -2\hat{b}\,(x_{2k} - \bar{x}_{2k}) ,$$

$$C = (x_{2k} - \bar{x}_{2k})^2 - t^2\hat{\sigma}^2 \; \frac{n+1}{n} \quad . \tag{3.9}$$

Thus, the probability statement in equation (3·5) is equivalent to

$$P[A(x_k' - \bar{x}_k')^2 + B(x_k' - \bar{x}_k') + C \leq 0] = 1 - \alpha \; . \tag{3·10}$$

A prediction interval for x_k' can now be obtained in terms of the roots of the quadratic equation (3·8) viz.,

$$(x_k' - \bar{x}_k') = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \; ; \tag{3·11}$$

the term under the radical sign simplifies to

$$B^2 - 4AC = 4t^2\hat{\sigma}^2 \left[\frac{\left(x_{2k} - \bar{x}_{2k}\right)^2}{SXX} + \frac{n+1}{n}\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)\right] .$$

and hence

$$x_k' = \bar{x}_k' + \frac{\hat{b}(x_{2k} - \bar{x}_{2k}) \pm t\hat{\sigma}\sqrt{\frac{(x_{2k}-\bar{x}_{2k})^2}{SXX} + \frac{n+1}{n}\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)}}{\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)} \tag{3·12}$$

Since $x_k' = (x_{2k-1} + x_{2k+1})/2$,

The unknown future value x_{2k+1} can be written in terms of the known value as

$$x_{2k+1} = D \pm E \text{ where}$$

$$D = 2\bar{x}_k' - x_{2k-1} + 2\hat{b}\,(x_{2k} - \bar{x}_{2k})/(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}) \qquad (3\cdot 13)$$

$$E = 2t\hat{\sigma}\sqrt{(x_{2k} - \bar{x}_{2k})^2/SXX + \frac{n+1}{n}(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX})} \Bigg/ \left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right) \qquad (3\cdot 14)$$

Thus a prediction interval for x_{2k+1} is given by

$$D - E \leq x_{2k+1} \leq D + E \ , \qquad (3\cdot 15)$$

where D and E are given by equations (3·13) and (3·14) respectively. The properties of this interval are discussed in section B.

B. PROPERTIES OF THE PREDICTION INTERVAL

The prediction interval in the previous section was obtained by solving

$$A(X_k' - \bar{X}_k')^2 + B(X_k' - \bar{X}_k') + C = 0$$

where the coefficients A, B and C are defined in (3·9). The roots of the equation and hence the form of the prediction interval will depend on the relative sizes of A, B and C.

The effect of these coefficients on the prediction interval will now be examined breaking in to four all inclusive cases.

<u>Case 1</u>. $A > 0$ and $B^2 - 4AC > 0$

$$A > 0 \Rightarrow \hat{b}^2 > \frac{t^2\hat{\sigma}^2}{SXX}$$

or $$\frac{\hat{b}^2 SXX}{t^2\hat{\sigma}^2} > 1 .$$

Let $$\frac{\hat{b}^2 SXX}{t^2\hat{\sigma}^2} = F \tag{3·16}$$

$B^2 > 4AC$ implies that

$$4\,\hat{b}^2\,(x_{2k} - \bar{x}_{2k})^2 > 4\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)(x_{2k} - \bar{x}_{2k})^2 -$$
$$4\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)\frac{t^2\hat{\sigma}^2}{n}(n+1)$$

or $$\left(\hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX}\right)\frac{t^2\hat{\sigma}^2}{n}(n+1) > -\frac{t^2\hat{\sigma}^2}{SXX}(x_{2k} - \bar{x}_{2k})^2$$

$$\frac{\hat{b}^2 SXX}{t^2\hat{\sigma}^2} - 1 > -\frac{n}{n+1}\frac{(x_{2k} - \bar{x}_{2k})^2}{t^2\hat{\sigma}^2}$$

or $$\frac{\hat{b}^2 SXX}{t^2\hat{\sigma}^2} > 1 - \frac{n}{n+1}\frac{(x_{2k} - \bar{x}_{2k})^2}{t^2\hat{\sigma}^2}$$

To satisfy the conditions $A > 0$ and $B^2 > 4AC$ it is necessary that

$$F > 1 \text{ and } F > 1 - \frac{n}{n+1}\frac{(x_{2k} - \bar{x}_{2k})^2}{t^2\hat{\sigma}^2} . \tag{3·16a}$$

Since the quantity $(x_{2k} - \bar{x}_{2k})^2/t^2\hat{\sigma}^2$ is always positive the conditions in equations (3·16a) are equivalent to the condition $F > 1$. In this case the prediction interval for x_{2k+1} will

be of the form

$$\left[(D - E) \; , \; (D + E)\right] .$$

This type of two-sided interval is referred to as TYPE 1, where D and E are as defined in (3·13) and (3·14) respectively.

<u>Case 2</u>. $B^2 - 4AC \geq 0$ and $A < 0$

$$A < 0 \Rightarrow \hat{b}^2 - \frac{t^2\hat{\sigma}^2}{SXX} < 0$$

$$\text{or} \quad \hat{b}^2 < \frac{t^2\hat{\sigma}^2}{SXX}$$

$$\text{or} \quad \frac{\hat{b}^2 SXX}{t^2\hat{\sigma}^2} < 1 .$$

$$\text{i.e.,} \quad A < 0 \Rightarrow F < 1. \tag{3·17}$$

$$B^2 - 4AC \geq 0 \Rightarrow F \geq 1 - \frac{n}{n+1} \frac{(x_{2k} - \bar{x}_{2k})^2}{t^2\hat{\sigma}^2} . \tag{3·18}$$

Thus, if $F < 1$ and $F \geq 1 - \frac{n}{n+1} \frac{(x_{2k} - \bar{x}_{2k})^2}{t^2\hat{\sigma}^2}$

the resulting interval will be of the form,

$$(-\infty, D + E] \; [D - E, \infty) \tag{3·19}$$

and will be called a TYPE 2 interval.

<u>Case 3</u>. If $B^2 - 4AC < 0$ and $A \neq 0$ the interval is called TYPE 3.

The roots of the quadratic are complex and the prediction interval must be taken as $(-\infty, +\infty)$. This situation occurs when $B^2 - 4AC < 0$.

$$B^2 - 4AC < 0 \Rightarrow F < 1 - \frac{n}{n+1} \frac{(x_{2k} - \bar{x}_{2k})^2}{t^2 \hat{\sigma}^2} \tag{3·20}$$

Thus a type 3 interval results when

$$F < 1 - \frac{n}{n+1} \frac{(x_{2k} - \bar{x}_{2k})^2}{t^2 \hat{\sigma}^2}$$

CASE 4. $A = 0$

$$A = 0 \Rightarrow \frac{\hat{b}^2 SXX}{t^2 \hat{\sigma}^2} = 1 . \tag{3.21}$$

Thus, when $F = 1$, the interval is of type 4 and is of the form $[G, \infty)$,

where $$G = 2\bar{x}_k' - x_{2k-1} - \hat{b}\ SXX\ (\frac{n+1}{n}) + \frac{x_{2k} - \bar{x}_{2k}}{\hat{b}} . \tag{3·22}$$

To summarize the results, the quadratic equation that leads to the required prediction interval is of the form

$$A(x_k' - \bar{x}_k')^2 + B(x_k' - \bar{x}_k') + C = 0,$$

and four different types of intervals can result depending on the value of

$$F = \frac{\hat{b}^2\ SXX}{t^2 \hat{\sigma}^2} .$$

The four cases are

i) when $F > 1$, the interval is of the form $[(D-E),(D+E)]$.

ii) when $F < 1$ and $F \geq 1-\frac{n}{n+1}\frac{(x_{2k}-\bar{x}_{2k})^2}{t^2\hat{\sigma}^2}$, the interval is of type 2 and is of the form $(-\infty, D+E],[D-E,\infty)$

iii) when $F < 1-\frac{n}{n+1}\frac{(x_{2k}-\bar{x}_{2k})^2}{t^2\hat{\sigma}^2}$, the interval is of type 3 and is of the form $(-\infty,+\infty)$

iv) when $F = 1$, the interval is of type 4 i.e., $[G,\infty)$.

At a given level of significance α, from the relation (3·5) and the conclusions drawn in this section, the following relation should hold:

$$\sum_{i=1}^{4} \Pr[\text{occurrence of type } i \text{ interval}] \times \Pr[i^{th} \text{ interval contains the predicted value}] = 1-\alpha \qquad (3\cdot 23)$$

The above conclusions are verified with simulated samples in chapter 5.

Section C examines the distribution of the quantity F defined in equation (3·16).

C. DISTRIBUTION OF 'F'

The occurrence of any particular type of interval depends on the value

$$F = \frac{\hat{b}^2\, SXX}{t^2\hat{\sigma}^2}$$

The distribution of F can be shown to be a noncentral F distribution as follows.

From (2·2), it can be shown that $\hat{b}$ is $N(b, \sigma^2/SXX)$

or $$\frac{(\hat{b}-b)}{\sigma/\sqrt{SXX}} \sim N(0,1) \quad ,$$

or $$\frac{(\hat{b}-b)^2}{\sigma^2/SXX} \sim \chi^2(1),$$

and $$\frac{\hat{b}^2 \cdot SXX}{\sigma^2} \sim \chi^2\,(1,\, \lambda) \quad . \qquad (3\cdot24)$$

where $$\lambda = \frac{b^2\; SXX}{\sigma^2} \quad ,$$

is the noncentrality parameter.

From (2·3) the distribution of

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \text{ is } \chi^2\,(n-2) \quad . \qquad (3\cdot25)$$

The ratio of (3·24) and (3·25) devided by their respective degress of freedom results in a noncentral F [5] statistic with degrees of freedom (1, n-2) and noncentrality parameter λ. That is

$$\frac{\hat{b}^2 \cdot SXX}{\hat{\sigma}^2} = t^2F \sim \mathbb{F}\,(1,\, n-2,\, \lambda) \quad , \qquad (3\cdot26)$$

and the noncentrality parameter

$$\lambda = \frac{b^2\; SXX}{\sigma^2} \quad .$$

Substituting for b from equation (2·5), the noncentrality parameter λ can be written as

$$\lambda = \frac{4\rho^2}{(1+\rho^2)^2} \cdot \frac{SXX}{\sigma^2} \qquad (3\cdot 27)$$

D. NUMERICAL EXAMPLE

The Model developed in section A is tested on the data obtained from Dow-Jones monthly averages to predict a future monthly average, based on the averages of the previous months. In most of the cases, the future value was contained in the prediction interval. To illustrate an example the data for the years 1966 and 67 are given below. The prediction intervals are computed with the help of the APL program PREDICT, whose listing is given on page 68.

1966			1967		
Month	Ended	D - J Ind	Month	Ended	D - J Ind
January	31	983.51	January	31	849.87
February	28	951.89	February	28	839.37
March	31	924.77	March	31	865.98
April	29	933.68	April	28	897.05
May	31	884.07	May	31	852.56
June	30	870.10	June	30	860.26
July	29	847.38	July	31	904.24
August	31	788.41	August	31	901.29
Sept	30	774.22	Sept	29	926.66
October	31	807.07	October	31	879.74
November	30	791.59	November	30	875.81
December	30	785.69	December	29	905.11

The above data are stored in the variable DMA. The program Predict is run with different sample sizes and the output is shown on pages 24 and 25.

Using 14 monthly averages from January '66 to February '67, prediction intervals for March '67 are computed and the exact value of 865.98 is contained in the interval.

Using given values up to April '67, the prediction intervals for May '67, also contain the true value. With a sample of size 18, the prediction intervals for the nineteenth month i.e., July '67 are computed and they contain the true value. Though the average of September '67 falls outside the interval, the average of November '67 is contained within the interval.

The reduction in the length of prediction interval with increase in sample size, is supported by the above example.

The data are tested for the required correlation structure and found to fulfill the requirements partially. The statistics of the data are

Mean	=	870.83
Standard deviation	=	55.0
Corr. Coefficient ρ	=	0.82

PREDICT 14↑DMA

THE F STATISTIC IS	=	4.132073877
THE LOWER PREDICTION LIMIT	=	727.1128824
THE UPPER PREDICTION LIMIT	=	1080.631201
THE LENGTH OF THE INTERVAL	=	353.5183188
TRUE VALUE OF THE AVERAGE	=	865.98

PREDICT 16↑DMA

THE F STATISTIC IS	=	5.489038103
THE LOWER PREDICTION LIMIT	=	598.3935849
THE UPPER PREDICTION LIMIT	=	900.5564036
THE LENGTH OF THE INTERVAL	=	302.1628187
TRUE VALUE OF THE AVERAGE	=	852.56

PREDICT 18↑DMA

THE F STATISTIC IS	=	5.130517543
THE LOWER PREDICTION LIMIT	=	708.7039
THE UPPER PREDICTION LIMIT	=	1009.101896
THE LENGTH OF THE INTERVAL	=	300.3979963
TRUE VALUE OF THE AVERAGE	=	904.24

PREDICT 20↑DMA

THE F STATISTIC IS	=	5.777973921
THE LOWER PREDICTION LIMIT	=	573.5929826
THE UPPER PREDICTION LIMIT	=	864.8705316
THE LENGTH OF THE INTERVAL	=	291.277549
TRUE VALUE OF THE AVERAGE	=	926.66

PREDICT 22↑DMA

THE F STATISTIC IS	=	6.994239226
THE LOWER PREDICTION LIMIT	=	633.1176654
THE UPPER PREDICTION LIMIT	=	899.7280383
THE LENGTH OF THE INTERVAL	=	266.6103729
TRUE VALUE OF THE AVERAGE	=	875.81

IV. SIMULATION OF SERIALLY CORRELATED SAMPLES

To study the effect of ρ, σ and n on the type of prediction interval obtained it is necessary to generate random variables with the required variance covariance matrix. These can be generated as linear combinations of standard normal random variates.

A. SIMULATION FROM STANDARD NORMAL SAMPLES

Suppose $Z_1, Z_2, \ldots, Z_n$ are random samples from $N(0, 1)$. To get serially correlated random variates $X_1, X_2, \ldots, X_n$, let

$$
\begin{aligned}
X_1 &= Z_1 \\
X_2 &= C_1 Z_1 + C_2 Z_2, \\
&\vdots \\
X_n &= C_1' Z_1 + C_2' Z_2 + \ldots\ldots + C_n' Z_n \quad . \qquad (4\cdot 1)
\end{aligned}
$$

The coefficients $C_1, \ldots, C_n$ are then determined by solving equations of the form,

$$\mathrm{Cov}\,(X_t, X_{t+k}) = \rho^k \text{ for all } t \text{ and } k \text{ from 1 to } n.$$

Consider the first equation

$$X_1 = Z_1$$

$$E(X_1) = E(Z_1) = 0$$

$$\mathrm{Var}\,(X_1) = \mathrm{Var}\,(Z_1) = 1 \qquad (4\cdot 2)$$

From the second equation

$$E(X_2) = E(C_1Z_1+C_2Z_2)$$

$$= C_1E(Z_1)+C_2\ E(Z_2) = 0 \qquad (4\cdot3)$$

since Z_1 and Z_2 are independent;

$$\text{Var}\ (X_2) = \text{Var}(C_1Z_1+ C_2Z_2)$$

$$= C_1^2\ \text{Var}\ Z_1 + C_2^2\ \text{var}\ Z_2\text{, i.e.,}$$

$$C_1{}^2 + C_2{}^2 = 1 \qquad (4\cdot4)$$

$$\text{Cov}(X_1,\ X_2) = E\ [(X_1 - E(X_1))(X_2 - E(X_2))]$$

$$= E\ [X_1 \cdot X_2]$$

$$= E\ [Z_1 \cdot (C_1Z_1+C_2Z_2)]$$

$$= E\ [C_1Z_1{}^2+C_2Z_1Z_2]\ \text{, i.e,}$$

$$C_1E(Z_1{}^2)+C_2E(Z_1Z_2) = \rho\ .$$

Since $Z_1 \sim N(0,\ 1)$, $Z_1{}^2 \sim \chi^2(1)$;

also since Z_1 and Z_2 are independent, the above equation can be simplified to

$$C_1 = \rho \qquad (4\cdot5)$$

From 4.4 by substitution we get $C_2 = \sqrt{1 - \rho^2}$

$$\text{or} \quad X_2 = \rho Z_1 + \sqrt{1 - \rho^2}\ \ Z_2 \qquad (4\cdot6)$$

$$\text{or} \quad X_2 = \rho X_1 + \sqrt{1 - \rho^2}\ \ Z_2 \qquad (4\cdot7)$$

Similarly it can be shown that

$$X_3 = \rho X_2 + \sqrt{1 - \rho^2} \quad Z_3 \tag{4·8}$$

The result can be generalised in the form

$$X_i = \rho X_{i-1} + \sqrt{1 - \rho^2} \quad Z_i \tag{4·9}$$

$$i = 2, \ldots, n.$$

or $$X_i = \rho^{i-1} Z_1 + \rho^{i-2}\sqrt{1 - \rho^2} \; Z_2 + \ldots + \sqrt{1 - \rho^2} \; Z_i \tag{4·10}$$

$$i = 1, \ldots, n.$$

The results obtained in (4·10) can be expressed in the following matrix notation:

$$\underset{n\text{x}1}{\underline{X}} = \underset{n\text{x}n}{\underline{A}} \; \underset{n\text{x}1}{\underline{Z}} \quad , \tag{4·11}$$

where $\underset{n\text{x}1}{\underline{X}} = (X_1, X_2, \ldots, X_n)^T$,

$\underset{n\text{x}1}{\underline{Z}} = (Z_1, Z_2, \ldots, Z_n)^T$,

and $\underline{A}$ is a nxn matrix of the form

$$\underline{A} = \begin{pmatrix} 1 & 0 & 0 \; \ldots\ldots\ldots \; 0 \\ \rho & \sqrt{1 - \rho^2} & 0 \; \ldots\ldots\ldots \; 0 \\ \rho^2 & \rho\sqrt{1 - \rho^2} & \sqrt{1 - \rho^2} \; \ldots\ldots \; 0 \\ \rho^{n-1} & \rho^{n-2}\sqrt{1 - \rho^2} & \ldots\ldots \sqrt{1 - \rho^2} \end{pmatrix} \tag{4·12}$$

If a random variable $\underset{n\text{x}1}{\underline{Z}}$ is distributed normally with mean $\underline{\mu}$ and covariance $\underline{V}$ and if $\underline{A}$ is an nxn matrix of rank n, then the vector

$$\underset{n \times 1}{\underline{Y}} = \underset{n \times n}{\underline{A}} \cdot \underset{n \times 1}{\underline{Z}}$$

is distributed as multivariate normal with mean vector $\underline{A}\ \underline{\mu}$ and covariance $\underline{A}\ \underline{V}\ \underline{A}^T$. As a result, the random variable $\underline{X}$ (4·11) is distributed as multivariate normal with the required covariance matrix $\underline{V}$. Here

$$\underline{Z} \sim N(\underline{0}, \underline{I}) ,$$

where $\underline{I}$ is an nxn identity matrix.

Hence, $\underline{X} \sim N(\underline{0}, \underline{A}\ \underline{I}\ \underline{A}^T)$.

We now show that $\underline{A}\ \underline{I}\ \underline{A}^T = \underline{A}\ \underline{A}^T$ is a matrix of the required form (for n = 4)

$$\underset{4 \times 4}{\underline{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho & \sqrt{1-\rho^2} & 0 & 0 \\ \rho 2 & \rho\sqrt{1-\rho^2} & \sqrt{1-\rho^2} & 0 \\ \rho^3 & \rho^2\sqrt{1-\rho^2} & \rho\sqrt{1-\rho^2} & \sqrt{1-\rho^2} \end{bmatrix}$$

$$\underset{4 \times 4}{\underline{A}^T} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ 0 & \sqrt{1-\rho^2} & \rho\sqrt{1-\rho^2} & \rho^2\sqrt{1-\rho^2} \\ 0 & 0 & \sqrt{1-\rho^2} & \rho\sqrt{1-\rho^2} \\ 0 & 0 & 0 & \sqrt{1-\rho^2} \end{bmatrix}$$

$$\underline{A}\,\underline{A}^T = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & \rho^2+1-\rho^2 & \rho^3+\rho(1-\rho^2) & \rho^4+\rho^2(1-\rho^2) \\ \rho^2 & \rho^3+\rho(1-\rho^2) & \rho^4+\rho^2(1-\rho^2)+(1-\rho^2) & \rho^5+\rho^3(1-\rho^2)+\rho(1-\rho^2) \\ \rho^3 & \rho^4+\rho^2(1-\rho^2) & \rho^5+\rho^3(1-\rho^2)+\rho(1-\rho^2) & \rho^6+(1-\rho^2)(\rho^4+\rho^2+1) \end{bmatrix}$$

which reduces to

$$\underline{A}\,\underline{A}^T = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

The above result can be generalized for any value of n.

B. TESTING FOR NORMALITY AND CORRELATION

The simulated samples are tested for normality and the correlation structure. The simulated samples are tested by standard Kolmogorov-Smirnov test. The results showed that samples do appear to be from the specified normal distribution.

The correlation between X_i and X_j is given by

$$R = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}\, X_i \cdot \mathrm{Var}\, X_j}} \qquad (4\cdot 13)$$

Confidence limits for correlation coefficient ρ are developed using the approximation.

$$Z = \tfrac{1}{2} \ln \frac{1+R}{1-R} \sim N\left(\tfrac{1}{2} \ln \frac{1+\rho}{1-\rho}, \frac{1}{n-3}\right)$$

A 95% confidence interval for ρ is given by

$$(\varepsilon^{2a}-1)/(\varepsilon^{2a}+1) < \rho < (\varepsilon^{2b}-1)/(\varepsilon^{2b}+1), \qquad (4\cdot14)$$

where

$$a = Z - \frac{1.96}{\sqrt{n-3}} \text{ and } b = Z + \frac{1.96}{\sqrt{n-3}}$$

The simulated samples are also tested for the required correlation structure by examining the confidence limits for correlation between (X_i, X_j) using equation (4·14). The results showed that the samples do appear to have the prescribed correlation structure.

V. RESULTS OF SIMULATION

In this chapter, using simulated samples, the probabilities of occurrence of various types of intervals and their dependence on the value of ρ, n and σ are studied. The program PROBS FORTRAN, whose listing is given on pages 63-66 computes the probabilities for the four types of intervals. For given values of ρ, n and σ the program generates one thousand samples and records the frequency of occurrence of each type of interval. In each case the program tests whether the predicted value is contained within the interval. Finally, for a given level of significance α the relation (3·23) is verified.

Choosing a level of significance $\alpha = 0.05$, the simulation was run for values of ρ ranging from 0.1 to 0.9, for sample sizes from 10 to 50 and for sigma varying from 1 to 5. For $\alpha = 0.1$ and 0.02 the program is executed to study the effect of the level of significance on the results.

The computer output is shown on pages 43 to 57. Tables I to VI show the variation of probabilities of types of intervals with change in α at different sample sizes, keeping α and σ constant. Tables VII to IX show the variation of probabilities of types of intervals with a change in ρ at values of $\alpha = 0.1$. Tables X to XII show the variation of probabilities of types of intervals with a change in ρ, keeping $\alpha = 0.02$.

In Tables XIII to XV, the variation of the probabilities of types of intervals are shown at a standard deviation (σ) of 3.0, whereas in Tables XVI to XVIII, the value of standard deviation is altered to 5.0. Tables XIX to XXVII depict the variation of the types of intervals with change in sample sizes at different values of ρ and σ, keeping α at 0.05. Finally, Tables XXVIII to XXX show the variation of probability of types of intervals with change in sigma at different values of ρ and n.

In all these tables the first column indicates the value of the parameter. Columns 2, 3 and 4 give the probabilities of occurrence of types of intervals. Columns 5 and 6 show the probability that the predicted value is contained in the interval. Since a type 3 interval is of the form $(-\infty,+\infty)$ the probability of x_{2k+1} to be contained in the interval is one. Finally column 7 verifies the relation (3.23). The attached graphs on pages 58 to 62 show the trend in the probability of occurence of the three types of intervals with changes in parameters like ρ, σ, n and α.

The outcome of the simulation confirmed the following theoretical conclusions:

i) The probability of type 1 interval increases with an increase in correlation coefficient ρ, all other parameters remaining constant.

ii) The probability of type 1 interval increases with an increase in sample size, keeping all other parameters constant.

iii) The probabilities of types 2 and 3 decrease with an increase in correlation coefficient or sample size.

iv) The probabilities for the four types of intervals do not change significantly with change in sigma.

v) The probability of a type 1 interval increases with an increase in level of significance α, all other parameters remaining constant.

Section 6 deals with the application to linear trend model.

VI. APPLICATION TO LINEAR TREND MODEL

In this chapter, the results of chapter 3 are extended to stochastic processes that follow a linear trend model.

Let $\{X_n\}$ be a discrete stochastic process satisfying

$$X_n = \mu_n + \varepsilon_n , \tag{6·1}$$

where
$$\mu_n = \alpha + \beta n , \tag{6·2}$$

and ε_n is a stationary process of normal variates that satisfy the following:

$$\begin{aligned} E\{\varepsilon_n\} &= 0 \\ Var\{\varepsilon_n\} &= \sigma^2 \\ Cor\{\varepsilon_n, \varepsilon_{n+k}\} &= \sigma^2 \rho^k \quad , \rho < 1 \quad . \end{aligned} \tag{6·3}$$

Such a process is called a model for linear trend. Krishnaih and Murthy [7] showed that for the above model the conditional distributions of X_{2k} given X_{2k-1} and X_{2k+1} are independent normal with

$$E\,[X_{2k} \mid X_{2k-1}, X_{2k+1}] = \beta_1 + \beta_2 K + \beta_3 X_k' \quad , \tag{6·4}$$

and
$$Var\,[X_{2k} \mid X_{2k-1}, X_{2k+1}] = \sigma^2(1-\rho^2)/(1+\rho^2) \quad , \tag{6·5}$$

where
$$X_k' = (X_{2k-1} + X_{2k+1})/2 \quad .$$

In section A, a prediction interval for X_{2k+1} is obtained knowing the values up to X_{2k} for the above model.

A. PREDICTION INTERVAL FOR X_{2k+1}

Using regression analysis, a prediction interval for X_{2k+1} can be obtained for the linear trend model, knowing the values up to X_{2k}. Equation (6·4) can be written as

$$X_{2k} = \beta_1 + \beta_2 k + \beta_3 X_k' + \varepsilon_k, \quad k = 1,2, \ldots, n,$$

or in matrix notation,

$$\underset{n\times 1}{\underline{X}}{}_{2k} = \underset{n\times 3}{\underline{X}} \; \underset{3\times 1}{\underline{B}} + \underset{n\times 1}{\underline{\varepsilon}} \quad , \quad \underline{\varepsilon} \sim N(\underline{0} \; , \sigma^2 \underline{I}) \; . \qquad (6\cdot 6)$$

The design matrix $\underset{n\times 3}{\underline{X}}$ is of the form:

$$\begin{bmatrix} 1 & 1 & (X_1+X_3)/2 \\ 1 & 2 & (X_3+X_5)/2 \\ 1 & 3 & (X_5+X_7)/2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & n & (X_{2n-1}+X_{2n+1})/2 \end{bmatrix}$$

and $\underset{3\times 1}{\underline{B}}$ is the column vector $(\beta_1, \beta_2, \beta_3)^T$

The estimator $\hat{\underline{B}}$ can be obtained from the normal equations

$$\underset{3\times n}{\underline{X}^T}\ \underset{n\times 3}{\underline{X}}\ \underset{3\times 1}{\hat{\underline{B}}} = \underset{3\times n}{\underline{X}^T}\ \underset{n\times 1}{\underline{X}}{}_{2k} ,$$

or $\quad \hat{\underline{B}} = \underline{S}^{-1}\underline{X}^T\underline{X}_{2k}$,

where $\quad \underline{S} = \underset{3\times n}{\underline{X}^T}\ \underset{n\times 3}{\underline{X}} = \begin{bmatrix} n & \Sigma n & \Sigma X'_n \\ \Sigma n & \Sigma n^2 & \Sigma n X'_n \\ \Sigma X'_n & \Sigma n X'_n & \Sigma X'^2_n \end{bmatrix} \qquad (6\cdot 7)$

Since $\underline{S}$ can be shown to be nonsingular and of rank 3, $\underline{S}^{-1}$ exists and

$$\hat{B} \sim N(\underline{B}, \sigma^2 \underline{S}^{-1}) \quad . \qquad (6\cdot 8)$$

For a given value of $\underset{1\times 3}{\underline{X}'}{}_k = [1\ k\ x'_k\ ]$,

the predicted value $\underset{1\times 1}{\hat{\underline{X}}}{}_{2k} = \underset{1\times 3}{\underline{X}'}{}_k\ \underset{3\times 1}{\hat{\underline{B}}}$,

and $\hat{\underline{X}}_{2k} \sim N\{\underline{X}'_k\ \underline{B}\ ,\ \sigma^2(\underline{X}'_k\ )\underline{S}^{-1}(\underline{X}'_k\ )^T\}$

It can be shown that the residual

$$x_{2k} - \hat{x}_{2k} \sim N\{0,\ \sigma^2(1+s^2)\}\ , \qquad (6\cdot 9)$$

where

$$\underset{1\times 1}{s^2} = \underset{1\times 3}{\underline{X}'}{}_k\ \underset{3\times 3}{\underline{S}^{-1}}(\underset{3\times 1}{\underline{X}'}{}_k\ )^T \quad .$$

As an example, let $\underline{X} = (x_1,\ x_2,\ \ldots, x_{10})^T$

Here K = 5 and n = 4.

The vector $\underset{4\times 1}{\underline{X}}{}_{2k} = (x_2,\ x_4,\ x_6,\ x_8)^T$,

and the design matrix $\underset{4\times 3}{\underline{X}} = \begin{bmatrix} 1 & 1 & (x_1+x_3)/2 \\ 1 & 2 & (x_3+x_5)/2 \\ 1 & 3 & (x_5+x_7)/2 \\ 1 & 4 & (x_7+x_9)/2 \end{bmatrix}$

$$\underline{X}_k' = [1\ 5\ (x_9+x_{11})/2] \quad .$$

The value of x_{11} is unknown. A prediction interval for x_{11} can be developed as shown below.

From equation (6·9)

$$\frac{x_{2k} - \hat{x}_{2k}}{\sigma\sqrt{1 + s^2}} \sim N\,(0,\ 1) \qquad (6{\cdot}10)$$

and

$$\frac{x_{2k} - \hat{x}_{2k}}{\sqrt{1 + s^2}} \bigg/ \frac{R_o}{\sqrt{n-3}} = t \ . \qquad (6{\cdot}11)$$

The quantity t has a student's t distribution with (n-3) degrees of freedom, where R_o^2 is the usual unbiased estimator of σ^2.

At a given level of significance α ,

$$P\left[\ \left|\frac{(x_{2k} - \hat{x}_{2k})\ \sqrt{n-3}}{R_o\ \sqrt{1 + s^2}}\right| \leqslant t_{\alpha/2}\right] = 1 - \alpha \qquad (6{\cdot}12)$$

The term in the brackets of the left hand side of equation (6·12) can be written as,

$$x_{2k} - \hat{x}_{2k} \leq R_o t_{\alpha/2} \sqrt{1+s^2}/\sqrt{n-3} \qquad (6\cdot13)$$

and

$$1 + s^2 = 1 + [1\ k\ x_k'][S^{-1}][1\ k\ x_k']^T.$$

The expression on the right hand side can be simplified to the form

$$\gamma_o + \gamma_1 x_k' + \gamma_2 x_k'^2 .$$

The coefficients γ_o, γ_1 and γ_2 can be computed since k and S^{-1} are known.

Squaring both sides of equation (6·13) and substituting for $\hat{x}_{2k}$ and s^2, it can be simplified to,

$$(x_{2k}-\hat{\beta}_1-\hat{\beta}_2 k-\hat{\beta}_3 x_k')^2 \leq \frac{t^2_{\alpha/2}R_o^2}{n-3}(\gamma_o+\gamma_1 x_k'+\gamma_2 x_k'^2)$$

or

$$(\hat{\beta}_3^2-\gamma_2\frac{t^2_{\alpha/2}R_o^2}{n-3})\, x_k'^2 + (2(x_{2k}-\hat{\beta}_1-\hat{\beta}_2 k)\hat{\beta}_3 - \frac{t^2_{\alpha/2}R_o^2\gamma_1}{n-3})\, x_k'$$

$$+ ((x_{2k}-\hat{\beta}_1-\hat{\beta}_2 k)^2 - \frac{t^2_{\alpha/2}R_o^2\gamma_o}{n-3}) \leq 0 .$$

The above inequality reduces to

$$A x_k'^2 + B x_k' + C \leq 0 , \qquad (6\cdot14)$$

where

$$A = \hat{\beta}_3^2 - \gamma_2 t^2_{\alpha/2}R_o^2/(n-3) ,$$

$$B = 2(x_{2k}-\hat{\beta}_1-\hat{\beta}_2 k)\hat{\beta}_3 - \frac{t^2_{\alpha/2}R_o^2\gamma_1}{n-3}$$

$$C = (x_{2k} - \hat{\beta}_1 - \hat{\beta}_2 k)^2 - \frac{t^2_{\alpha/2} R^2_o \gamma_o}{n-3} \qquad (6\cdot15)$$

The quadratic equation (6·14) can be solved for x'_{ko} using the formula.

$$x'_k = \frac{-B \pm \sqrt{B^2-4AC}}{2A} \quad \text{if } A \neq 0 .$$

Since $x'_k = \frac{1}{2}(x_{2k-1}+x_{2k+1})$, the prediction interval for x_{2k+1} can be written as

$$D - E < x_{2k+1} < D + E \qquad (6\cdot16)$$

where

$$D = -x_{2k-1} - B/A$$

$$E = \sqrt{(B^2-4AC)}/A \qquad (6\cdot17)$$

Note: If A=0, then no interval exists; also when $B^2 < 4AC$ E becomes imaginary and the interval is taken to be $(-\infty,+\infty)$.

In section A, a numerical example is given to verify the model developed here, with the help of simulated samples.

B. NUMERICAL EXAMPLE

In order to compute the prediction intervals for a future value for samples satisfying a linear trend model an APL program LINTREND has been developed. The listing of this program is given on page 67. Error terms having the correlation structure (6·3) are given as input and for given values of α and β, the program develops the vectors $\underline{X}_{2k}$ and $\underline{X}'_k$.

For a sample of size 18, the output of the above program is given below.

$\alpha = 1.0$, $\beta = 0.5$, $\rho = 0.2$, $N = 18$, $n = 8$.

$$\underset{8\times 1}{X}'_{k} = (2.094,\ 3.37,\ 4.056,\ 3.7595,\ 5.984,\ 7.848,\ 7.821,\ 8.388)^T$$

$$\underset{8\times 1}{X}_{2k} = (2.94,\ 3.983,\ 4.913,\ 2.685,\ 5.376,\ 7.849,\ 9.017,\ 8.416)^T$$

The design matrix $\underset{8\times 3}{X} = \begin{bmatrix} 1 & 1 & 2.094 \\ 1 & 2 & 3.37 \\ 1 & 3 & 4.056 \\ 1 & 4 & 3.759 \\ 1 & 5 & 5.984 \\ 1 & 6 & 7.848 \\ 1 & 7 & 7.821 \\ 1 & 8 & 8.388 \end{bmatrix}$

For the model (6·4),

$$\underset{3\times 1}{\hat{B}} = (-0.147,\ -0.597,\ 1.566)^T$$

The coefficients of the quadratic equation (6·14)

$$A(X_k')^2 + B(X_k') + C = 0 ,$$

are given by

$A = 1.443$

$B = 35.801$

$C = -72.382$

and the values of D and E are

$D = 15.955$

$E = 28.561$

from the equation (6·17) ,

the lower limit = - 12.606

the upper limit = 44.516 .

The true value of X_{19} (10.515) is contained in the interval.

I. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 14
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.193	0.066	0.741	0.917	0.576	0.956
0.300	0.342	0.080	0.578	0.915	0.775	0.953
0.500	0.269	0.098	0.633	0.883	0.786	0.949
0.700	0.369	0.119	0.512	0.902	0.874	0.949
0.900	0.539	0.128	0.333	0.939	0.930	0.958

II. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 22
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.492	0.048	0.460	0.935	0.438	0.941
0.300	0.516	0.077	0.407	0.924	0.714	0.939
0.500	0.483	0.110	0.407	0.909	0.818	0.936
0.700	0.791	0.071	0.138	0.934	0.873	0.939
0.900	0.902	0.050	0.048	0.947	0.920	0.948

CON REG = CONFIDENCE REGION

III. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 30
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.554	0.036	0.410	0.960	0.722	0.968
0.300	0.861	0.033	0.106	0.940	0.758	0.940
0.500	0.874	0.044	0.082	0.952	0.841	0.951
0.700	0.914	0.028	0.058	0.951	0.929	0.953
0.900	0.979	0.008	0.013	0.952	1.000	0.953

IV. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 34
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.585	0.041	0.374	0.959	0.610	0.960
0.300	0.683	0.066	0.251	0.943	0.773	0.946
0.500	0.718	0.077	0.205	0.937	0.766	0.937
0.700	0.950	0.013	0.037	0.942	0.846	0.943
0.900	0.992	0.004	0.004	0.948	1.000	0.948

V. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 38
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.703	0.030	0.267	0.949	0.467	0.948
0.300	0.809	0.036	0.155	0.954	0.806	0.956
0.500	0.847	0.041	0.112	0.956	0.854	0.957
0.700	0.966	0.015	0.019	0.959	0.933	0.959
0.900	0.994	0.004	0.002	0.957	1.000	0.957

VI. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 42
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.573	0.047	0.380	0.932	0.723	0.948
0.300	0.719	0.044	0.237	0.947	0.773	0.952
0.500	0.853	0.047	0.100	0.943	0.894	0.946
0.700	0.980	0.007	0.013	0.948	0.857	0.948
0.900	0.998	0.001	0.001	0.947	1.000	0.947

VII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 14
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.10

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.269	0.111	0.620	0.870	0.468	0.906
0.300	0.441	0.108	0.451	0.846	0.667	0.896
0.500	0.380	0.138	0.482	0.858	0.732	0.909
0.700	0.501	0.138	0.361	0.858	0.790	0.900
0.900	0.666	0.114	0.220	0.883	0.868	0.907

VIII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 22
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.10

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.524	0.077	0.399	0.876	0.468	0.894
0.300	0.599	0.101	0.300	0.873	0.564	0.880
0.500	0.617	0.122	0.261	0.861	0.754	0.884
0.700	0.853	0.066	0.081	0.885	0.773	0.887
0.900	0.941	0.023	0.031	0.891	0.786	0.891

IX. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 30
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.10

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.595	0.070	0.335	0.899	0.571	0.910
0.300	0.906	0.025	0.069	0.891	0.560	0.890
0.500	0.933	0.024	0.043	0.902	0.875	0.906
0.700	0.951	0.022	0.027	0.902	0.909	0.905
0.900	0.989	0.006	0.005	0.897	1.000	0.898

X. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 14
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.02

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.160	0.028	0.812	0.969	0.571	0.983
0.300	0.254	0.042	0.704	0.949	0.738	0.976
0.500	0.169	0.063	0.768	0.929	0.825	0.977
0.700	0.209	0.088	0.703	0.933	0.932	0.980
0.900	0.381	0.121	0.498	0.961	0.975	0.982

XI. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 22
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.02

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.464	0.026	0.510	0.970	0.731	0.979
0.300	0.427	0.054	0.519	0.967	0.870	0.979
0.500	0.338	0.076	0.586	0.938	0.987	0.978
0.700	0.691	0.073	0.236	0.965	0.973	0.974
0.900	0.825	0.073	0.102	0.979	0.986	0.982

XII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 30
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.02

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.525	0.019	0.456	0.983	0.737	0.986
0.300	0.830	0.022	0.148	0.972	0.773	0.972
0.500	0.810	0.032	0.158	0.978	0.938	0.980
0.700	0.846	0.051	0.103	0.979	0.961	0.980
0.900	0.961	0.015	0.024	0.977	1.000	0.978

XIII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 22
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.067	0.078	0.855	0.851	0.500	0.951
0.300	0.179	0.125	0.696	0.855	0.704	0.937
0.500	0.445	0.117	0.438	0.899	0.829	0.935
0.700	0.736	0.083	0.176	0.938	0.898	0.945
0.900	0.901	0.051	0.048	0.948	0.941	0.950

XIV. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 26
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.652	0.043	0.305	0.936	0.651	0.943
0.300	0.255	0.106	0.639	0.929	0.698	0.950
0.500	0.561	0.110	0.329	0.938	0.855	0.949
0.700	0.841	0.062	0.097	0.941	0.903	0.944
0.900	0.962	0.016	0.022	0.946	0.938	0.947

XV.VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 30
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.075	0.071	0.854	0.880	0.620	0.964
0.300	0.374	0.111	0.515	0.922	0.721	0.940
0.500	0.674	0.091	0.235	0.942	0.857	0.948
0.700	0.897	0.034	0.069	0.950	0.912	0.952
0.900	0.979	0.007	0.014	0.952	1.000	0.953

XVI.VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 22
SIGMA = 5.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.067	0.078	0.855	0.851	0.500	0.951
0.300	0.182	0.127	0.691	0.863	0.717	0.939
0.500	0.444	0.115	0.441	0.901	0.817	0.935
0.700	0.716	0.102	0.182	0.936	0.892	0.943
0.900	0.901	0.052	0.047	0.948	0.942	0.950

XVII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 26
SIGMA = 5.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.193	0.097	0.710	0.896	0.619	0.943
0.300	0.255	0.107	0.638	0.929	0.701	0.950
0.500	0.560	0.111	0.329	0.938	0.856	0.949
0.700	0.841	0.062	0.097	0.941	0.903	0.944
0.900	0.962	0.016	0.022	0.947	0.938	0.948

XVIII. VARIATION OF TYPES OF INTERVALS WITH ROW

SAMPLE SIZE = 30
SIGMA = 5.0
LEVEL OF SIGNIFICANCE = 0.05

ROW	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
0.100	0.075	0.075	0.850	0.893	0.640	0.965
0.300	0.350	0.110	0.540	0.920	0.709	0.940
0.500	0.664	0.097	0.239	0.950	0.845	0.952
0.700	0.897	0.034	0.069	0.950	0.912	0.952
0.900	0.979	0.007	0.014	0.952	1.000	0.953

XIX.VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.300
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.300	0.040	0.660	0.867	0.600	0.944
6	0.342	0.080	0.578	0.915	0.775	0.953
8	0.587	0.057	0.356	0.942	0.737	0.951
10	0.516	0.077	0.407	0.924	0.714	0.939
12	0.259	0.106	0.635	0.946	0.698	0.954
14	0.861	0.033	0.106	0.940	0.758	0.940

XX. VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.500
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.232	0.049	0.719	0.853	0.694	0.951
6	0.269	0.098	0.633	0.888	0.786	0.949
8	0.540	0.083	0.377	0.935	0.819	0.950
10	0.483	0.110	0.407	0.909	0.818	0.936
12	0.561	0.109	0.330	0.941	0.853	0.951
14	0.874	0.044	0.082	0.952	0.841	0.951

XXI. VARIATION OF TYPES OF INTERVALS WITH N

CORR. COEFF. = 0.700
SIGMA = 1.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.308	0.072	0.620	0.896	0.861	0.958
6	0.369	0.119	0.512	0.902	0.874	0.949
8	0.586	0.095	0.319	0.920	0.884	0.942
10	0.791	0.071	0.138	0.934	0.873	0.939
12	0.840	0.065	0.095	0.940	0.908	0.944
14	0.914	0.028	0.058	0.951	0.929	0.953

XXII. VARIATION OF TYPES OF INTERVALS WITH N

CORR. COEFF. = 0.300
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.067	0.050	0.883	0.642	0.580	0.955
6	0.090	0.093	0.817	0.689	0.710	0.945
8	0.262	0.085	0.653	0.924	0.718	0.956
10	0.179	0.125	0.696	0.855	0.704	0.937
12	0.255	0.106	0.639	0.929	0.698	0.950
14	0.374	0.111	0.515	0.922	0.721	0.940

XXIII.VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.500
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.108	0.068	0.824	0.713	0.750	0.952
6	0.226	0.105	0.669	0.863	0.781	0.946
8	0.376	0.118	0.506	0.923	0.822	0.950
10	0.445	0.117	0.438	0.899	0.829	0.935
12	0.561	0.110	0.329	0.938	0.855	0.949
14	0.674	0.091	0.235	0.942	0.857	0.948

XXIV. VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.700
SIGMA = 3.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.216	0.077	0.707	0.829	0.870	0.953
6	0.354	0.122	0.524	0.898	0.885	0.950
8	0.557	0.118	0.325	0.926	0.864	0.943
10	0.736	0.088	0.176	0.938	0.898	0.945
12	0.841	0.062	0.097	0.941	0.903	0.944
14	0.897	0.034	0.069	0.950	0.912	0.952

XXV. VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.300
SIGMA = 5.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.068	0.051	0.881	0.632	0.588	0.954
6	0.088	0.094	0.818	0.682	0.713	0.945
8	0.138	0.093	0.769	0.884	0.710	0.957
10	0.182	0.127	0.691	0.863	0.717	0.939
12	0.255	0.107	0.638	0.929	0.701	0.950
14	0.350	0.110	0.540	0.920	0.709	0.940

XXVI. VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.500
SIGMA = 5.0
LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.107	0.069	0.824	0.701	0.739	0.950
6	0.215	0.109	0.676	0.860	0.761	0.944
8	0.363	0.126	0.511	0.928	0.810	0.950
10	0.444	0.115	0.441	0.901	0.817	0.935
12	0.560	0.111	0.329	0.938	0.856	0.949
14	0.664	0.097	0.239	0.950	0.845	0.952

XXVII.VARIATION OF TYPES OF INTERVALS WITH N

CORR.COEFF. = 0.900

SIGMA = 5.0

LEVEL OF SIGNIFICANCE = 0.05

N	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
4	0.237	0.103	0.660	0.873	0.922	0.962
6	0.537	0.129	0.334	0.937	0.930	0.957
8	0.777	0.080	0.143	0.943	0.888	0.947
10	0.901	0.052	0.047	0.948	0.942	0.950
12	0.962	0.016	0.022	0.947	0.938	0.948
14	0.979	0.007	0.014	0.952	1.000	0.953

XXVIII.VARIATION OF INTERVALS WITH SIGMA

SAMPLE SIZE = 14

CORR.COEFF. = 0.700

LEVEL OF SIGNIFICANCE = 0.05

SIGMA	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
1	0.369	0.119	0.512	0.902	0.874	0.949
2	0.352	0.123	0.525	0.901	0.886	0.951
3	0.354	0.122	0.524	0.898	0.885	0.950
4	0.354	0.121	0.525	0.898	0.884	0.950
5	0.353	0.122	0.525	0.898	0.885	0.950

XXIX. VARIATION OF INTERVALS WITH SIGMA

SAMPLE SIZE = 18

CORR.COEFF. = 0.500

LEVEL OF SIGNIFICANCE = 0.05

SIGMA	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
1	0.540	0.083	0.377	0.935	0.819	0.950
2	0.396	0.116	0.488	0.919	0.819	0.947
3	0.376	0.118	0.506	0.923	0.822	0.950
4	0.372	0.118	0.510	0.922	0.822	0.950
5	0.363	0.126	0.511	0.928	0.810	0.950

XXX.VARIATION OF INTERVALS WITH SIGMA

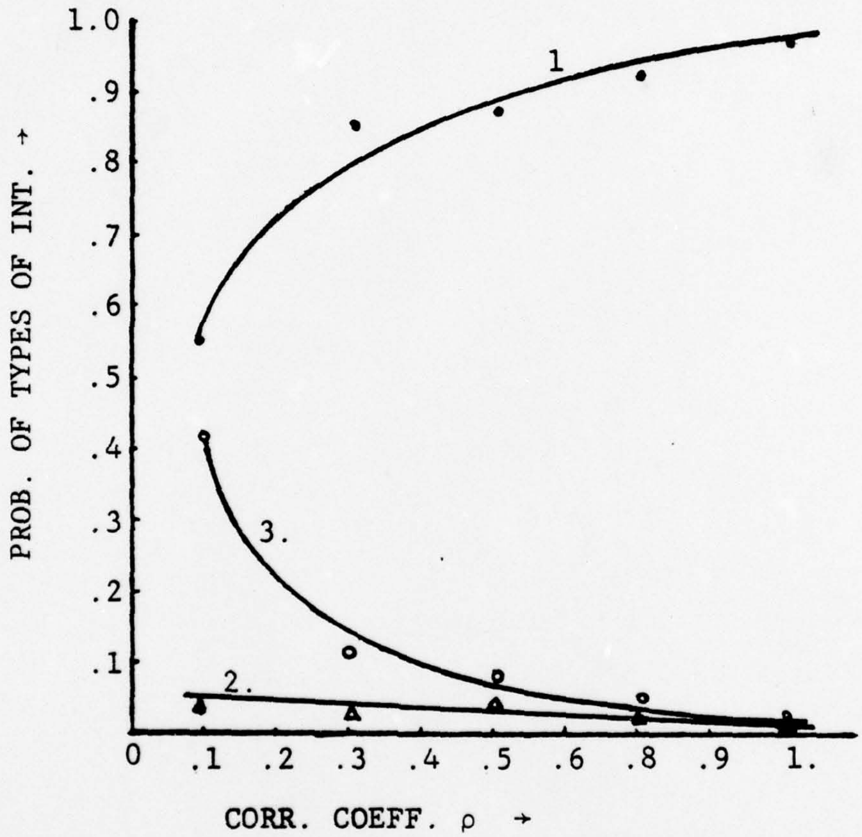
SAMPLE SIZE =22

CORR.COEFF. = 0.500

LEVEL OF SIGNIFICANCE = 0.05

SIGMA	TYPE 1	TYPE 2	TYPE 3	PROB 1	PROB 2	CON REG
1	0.483	0.110	0.407	0.909	0.818	0.936
2	0.446	0.115	0.439	0.901	0.826	0.936
3	0.445	0.117	0.438	0.899	0.828	0.935
4	0.446	0.115	0.439	0.899	0.826	0.935
5	0.444	0.115	0.441	0.901	0.817	0.935

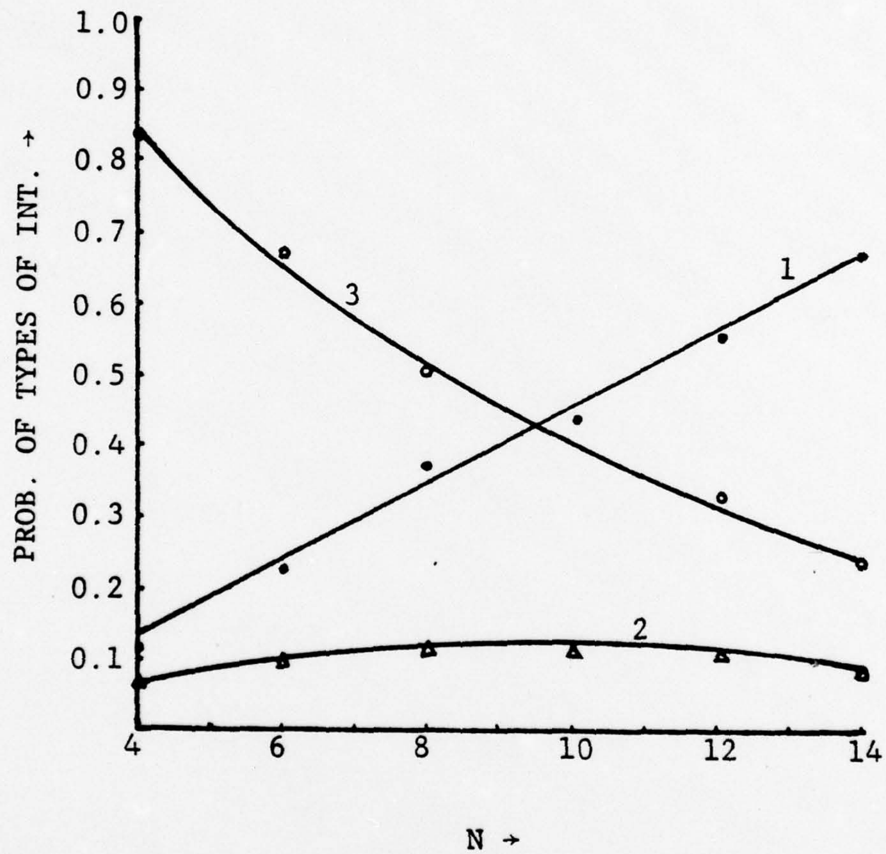
1. TYPES OF INTERVALS VS ρ



SAMPLE SIZE	=	30
SIGMA	=	1.0
SIGNIFICANCE LEVEL	=	0.05

1, 2 and 3 on the curves indicate the types of Intervals.

2. TYPES OF INTERVALS VS N

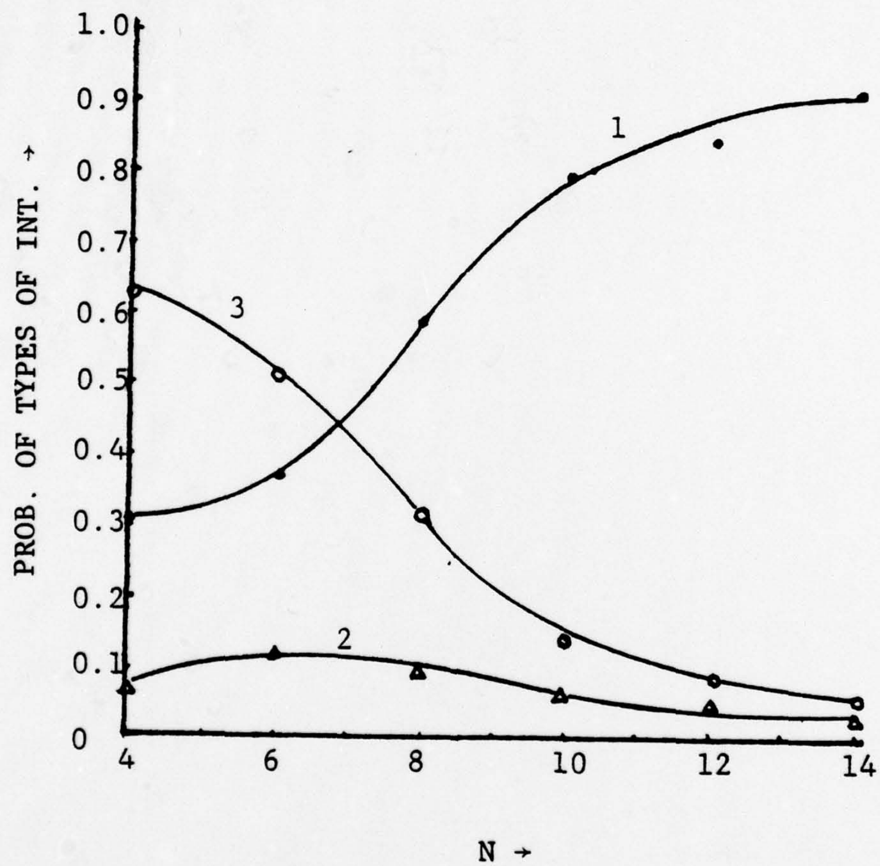


CORR. COEFF = 0.5

SIGMA = 3.0

SIGNIFICANCE LEVEL = 0.05

3. TYPES OF INTERVALS VS N

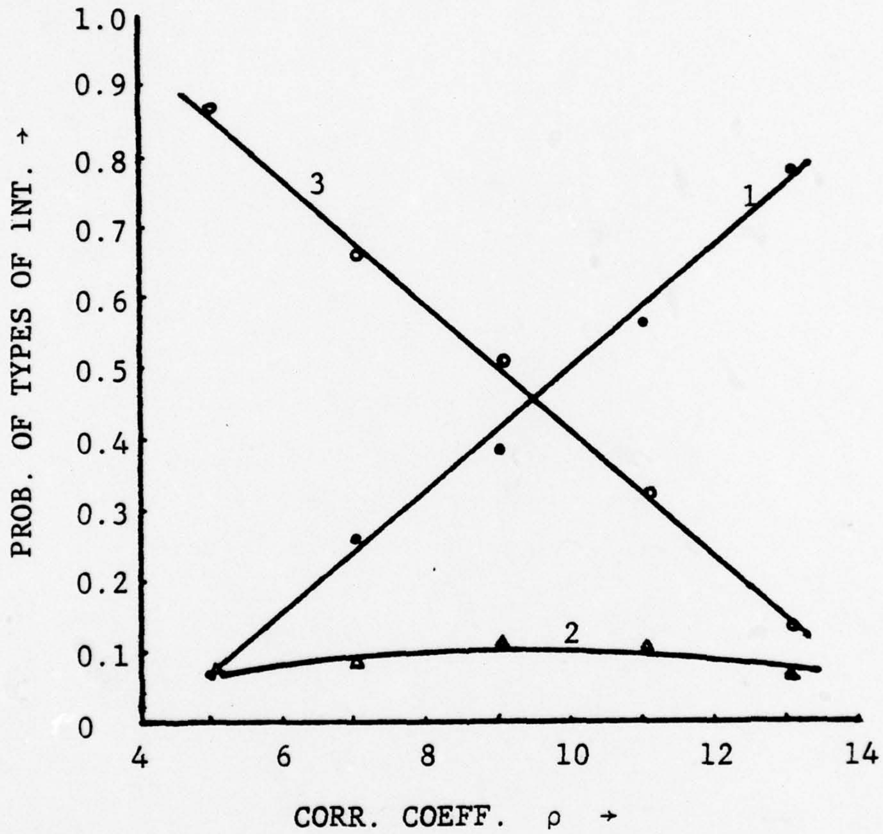


CORR. COEFF. = 0.7

SIGMA = 1.0

SIGNIFICANCE LEVEL = 0.05

4. TYPES OF INTERVALS VS ρ

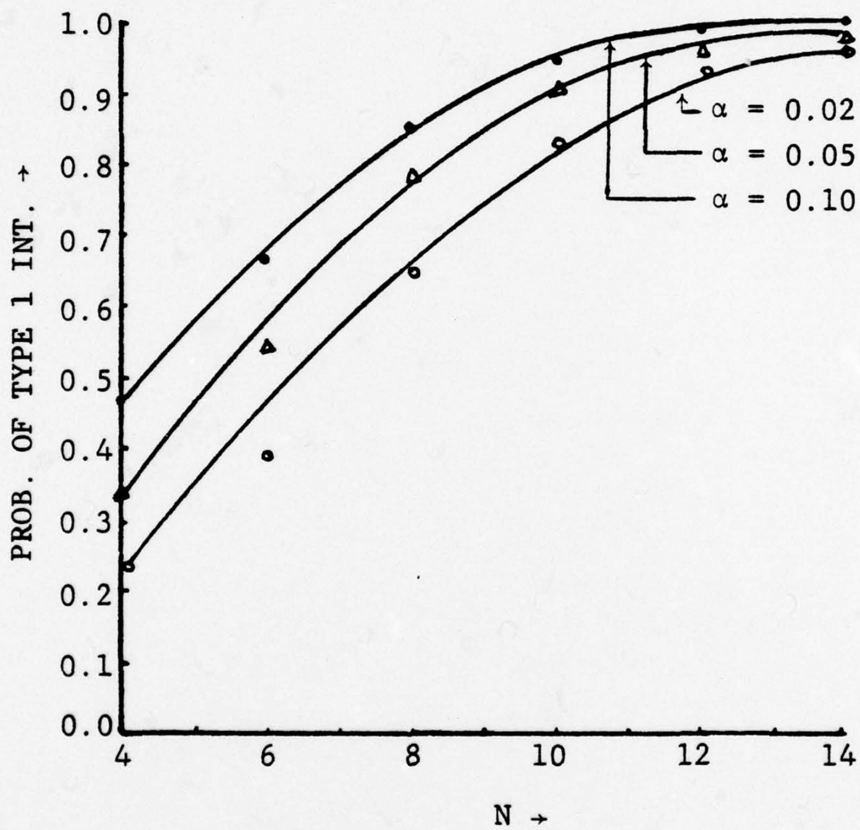


SAMPLE SIZE = 8

SIGMA = 3.0

SIGNIFICANCE LEVEL = 0.05

5. PROB. OF TYPE 1 INT. VS N
(VARIATION WITH α)



CORR. COEFF. = 0.9

SIGMA = 1.0

```
C     PROGRAM TO CALCULATE PROBABILITY OF TYPE 1,TYPE 2,TYPE
C     3,TYPE 4 INTERVALS FOR SIMULATED SAMPLES.
C     PROGRAMMER T S MURTHY   SEP 1979.
C     ************************************************************
      DIMENSION Z(55),X(55),V(100,55),XI(50),YI(50),S(55),
     1IC(5),IR(5),IPC(5),STAT(10,7),VS(10,10,7),IVS(10)
      CALL OVFLOW
      INDEX=1
      SIGMA=1.0
1     READ(5,2) K,T
      WRITE (6,2)K,T
2     FORMAT( 1X,I2,2X F5.3)
      IF(K .EQ. 0 ) GO TO 460
      ROW=0.10
      DO 300 IB=1,5
      DO 10 J=1,5
      B=(1-ROW**2 )**.5
C     SIMULATION OF SAMPLES
      ISEED=12345
10    IPC(J)=0
      DO 250 IA=1,10
      DO 50 M=1,100
      CALL SNORM(ISEED,Z,K)
      X(1)=SIGMA*Z(1)
      DO 30 J=2,K
30    X(J)=ROW*X(J-1)+B*SIGMA*Z(J)
      DO 40 L=1,K
40    V(M,L)=X(L)
50    CONTINUE
      DO 60 II=1,5
60    IR(II)=0
      DO 200 I=1,100
      DO 70 J=1,K
70    S(J)=V(I,J)
      K=K-5
      KK=K/2
```

```
      DO 80 L=1,KK
80    YI(L)=S(2*L)
      N=KK-1
      DO 90 LL=1,N
90    XI(LL)=(S(2*LL-1)+S(2*LL+1))/2.0
      XSUM=0.0
      YSUM=0.0
      SXX=0.0
      SXY=0.0
      SYY=0.0
      DO 100 KL=1,5
100   IC(KL)=0
      DO 110 M=1,N
      YSUM=YSUM+YI(M)
      XSUM=XSUM+XI(M)
110    CONTINUE
      XB=XSUM/N
      YB=YSUM/N
      DO 120 M=1,N
      SXX=SXX+(XI(M)-XB)**2
      SYY=SYY+(YI(M)-YB)**2
      SXY=SXY+(XI(M)-XB)*(YI(M)-YB)
120    CONTINUE
      VRES=(SYY-((SXY**2 )/SXX))/(N-2)
      BH=SXY/SXX
      AH=YB-BH*XB
      SS=(T**2 )*VRES
      A=(BH**2 )-SS/SXX
      P=S(2*KK)-YB
      B=-(2*BH*P)
      C=(P**2)-(SS*(N+1))/N
      F=(BH**2 )*SXX/SS
      SSQ=1.0-N*(P**2)/(SS*(N+1))
      IF (F .LT. SSQ) GO TO 150
      IF(F .EQ.1.0) GO TO 500
      D=(2.*XB)-S(K-1)+2.*BH*P/A
```

```
      E=(2./A)*(((SS*(P**2)/SXX)+(N+1)*A*SS/N)**.5)
      PIL=D-E
      PIR=D+E
      PVAL=S(K+1)
      IF(F .GT. 1.0) GO TO 130
      IC(2)=IC(2)+1
      IF(PVAL.LE.PIR .OR. PVAL.GE.PIL) IC(5)=IC(5)+1
      GO TO 160
130   IC(1)=IC(1)+1
      IF(PIL.LE.PVAL.AND.PIR.GE.PVAL) IC(4)=IC(4)+1
      GO TO 160
150   IC(3)=IC(3)+1
160   DO 170 J=1,5
170   IR(J)=IR(J)+IC(J)
      K=K+5
200   CONTINUE
      DO 220 J=1,5
220   IPC(J)=IPC(J)+IR(J)
250   CONTINUE
C     PRINT STATISTICS
      STAT(IB,1)=ROW
      STAT(IB,2)=IPC(1)/1000.0
      STAT(IB,3)=IPC(2)/1000.0
      STAT(IB,4)=IPC(3)/1000.0
      IF(STAT(IB,2).EQ.0.0) GO TO 275
      STAT(IB,5)=IPC(4)/(STAT(IB,2)*1000.0)
      GO TO 280
275    STAT(IB,5)=IPC(4)
280   IF(STAT(IB,3) .EQ. 0.0) GO TO 285
      STAT(IB,6)=IPC(5)/(STAT(IB,3)*1000.0)
      GO TO 290
285   STAT(IB,6)=IPC(5)
290   SIG=STAT(IB,2)*STAT(IB,5)+STAT(IB,3)*STAT(IB,6)
     1 +STAT(IB,4)
      STAT(IB,7)=SIG
      ROW=ROW+0.2
```

```
300   CONTINUE
      DO 305 IQ=1,5
      DO 305 JQ=1,7
      VS(INDEX,IQ,JQ)=STAT(IQ,JQ)
305   CONTINUE
      IVS(INDEX)=N
      INDEX=INDEX+1
      ISZ=K-5
      WRITE(6,310)ISZ,SIGMA,N
      WRITE (6,325)
      WRITE(6,350)((STAT(K,L),L=1,7),K=1,5)
310   FORMAT(1X,'   SAMPLE SIZE = ',I5,' SIGMA = ',F5.0,   '
     1N = ',I5,/)
325   FORMAT(1X,'    ROW      TYPE 1     TYPE 2     TYPE 3
     1PROB.1    PROB.2    CON.REG     ',/,70('-'))
350   FORMAT(  7(F8.3,2X),/       )
      WRITE(6,485)
      GO TO 1
460   CORR=0.1
      INDEX=INDEX-1
      DO 470 J=1,5
      WRITE (6,472) CORR,SIGMA
      WRITE(6,475)
       DO 471 I=1,INDEX
      WRITE(6,480)(IVS(I),(VS(I,J,K),K=2,7))
471   CONTINUE
      CORR=CORR+0.2
470   CONTINUE
472   FORMAT(  ' CORR.COEFF. = ',F5.3,' SIGMA = ',F5.0,//)
475   FORMAT(' SAMPLE SIZE TYPE 1     TYPE 2     TYPE 3     PRO
     1B.1    PROB.2    CON.REG ',/,70('-'))
480   FORMAT(I5,5X,6(F8.3,2X),/)
500   STOP
      END
```

APL PROGRAM LINTREND

```
     ∇ LINTREND X
[1]    X←1+(X+0.5×(⍳N←⍴X))
[2]    XF←X[N-1]
[3]    X←(N-2)↑X
[4]    K←(⍴X)×0.5
[5]    XI←⍳0
[6]    M←1
[7]    →11×⍳K<M←M+1
[8]    X1←(X[((2×M)-3)]+X[((2×M)-1)])×0.5
[9]    XI←XI,X1
[10]   →7
[11]   YI←X[2×⍳K]
[12]   N←K-1
[13]   YI←N↑YI
[14]   ' THE VECTOR XI IS ';XI
[15]   ' THE VECTOR YI IS ';YI
[16]   ' THE VALUE OF XF IS ';XF
[17]   C1←N⍴1
[18]   C2←⍳N
[19]   C3←XI
[20]   DMT←(3,N)⍴C1,C2,C3
[21]   DM←⍉DMT
[22]   S←DMT+.×DM
[23]   SI←⌹S
[24]   BH←(SI+.×DMT)+.×YI
[25]   SHS←(+/((YI)*2))-((⍉YI+.×DM)+.×BH)
[26]   A0←SI[1;1]+(N+1)×((2×SI[1;2])+(N+1)×SI[2;2])
[27]   A0←A0+1
[28]   A1←(2×SI[1;3])+2×SI[2;3]×(N+1)
[29]   A2←SI[3;3]
[30]   CF←4×SHS÷(N-3)
[31]   A←(BH[3]*2)-(A2×CF)
[32]   B←(2×BH[3]×W←(X[2×K]-BH[1]-BH[2]×K))-A1×CF
[33]   C←(W*2)-A0×CF
[34]   →IM×⍳(B*2)≤4×A×C
[35]   E←((B*2)-4×A×C)*0.5
[36]   E←E÷A
[37]   D←¯1×X[(2×K)-1]-B÷A
[38]   ' THE  LEFT LIMIT IS ';D-E
[39]   ' THE RIGHT LIMIT IS ';D+E
[40]   →0
[41] IM:' NO INTERVAL EXISTS '
     ∇
```

APL PROGRAM PREDICT

```
      ∇ PREDICT X;A;B;C;D;E;F;G;N;P;SSQ;SSX;SXX;SXY;SYY;XB;YB
[1]     ⍝ X IS ASSUMED TO BE A VECTOR OF EVEN NO OF ENTRIES
[2]     K←(⍴X)×0.5
[3]     XI←⍳0
[4]     M←1
[5]     →9×⍳K<M←M+1
[6]     X1←(X[((2×M)-3)]+X[((2×M)-1)])×0.5
[7]     XI←XI,X1
[8]     →5
[9]     YI←X[2×⍳K]
[10]    N←K-1
[11]    YI←N↑YI
[12]    YB←(+/YI)÷N
[13]    XB←(+/XI)÷N
[14]    SXY←+/((XI-XB)×(YI-YB))
[15]    SXX←+/((XI-XB)*2)
[16]    SYY←+/((YI-YB)*2)
[17]    BH←SXY÷SXX
[18]    AH←YB-BH×XB
[19]    VRES←(SYY+(¯1×((SXY*2)÷SXX)))÷N-2
[20]    G← 12.706 4.303 3.182 2.776 2.571 2.447 2.365 2.306
[21]    G←G, 2.201 2.179 2.16 2.145 2.131 2.12 2.11 2.101 2.09
[22]    G←G, 2.08 2.074 2.069 2.064 2.06 2.056 2.052 2.048 2
[23]    G←G, 2.041 2.04 2.04 2.04 2.035 2.03 2.03 2.03
[24]    G←G, 2.025 2.025 2.02 2.02 2.02 2.02 2.015 2.01 2.01
[25]    T←G[N-2]
[26]    A←(BH*2)-((÷SXX)×S←((T*2)×VRES))
[27]    B←¯1×(2×BH×P←(X[2×K]-YB))
[28]    C←(P*2)-S×((N+1)÷N)
[29]    F←(BH*2)×SXX÷S
[30]    ' THE F STATISTIC IS     ';F
[31]    SSQ←1-(N÷(N+1))×(P*2)÷S
[32]    →T1×⍳F>1
[33]    →T2×⍳F=1
[34]    →T3×⍳F<SSQ
[35]    ' THE INTERVAL IS OF TYPE 2 '
[36]    →0
[37]  T1:→⍳0
[38]    D←(2×XB)-X[(2×K)-1]+2×BH×P÷A
[39]    E←(÷A)×2×(S×(((P*2)÷SXX)+((N+1)÷N)×A))*0.5
[40]    ' THE LEFT LIMIT IS      ';D-E
[41]    ' THE RIGHT LIMIT IS     ';D+E
[42]    ' THE LENGTH OF INT IS   ';|(2×E)
[43]    →0
[44]  T2:' THE INT IS OF TYPE 4'
[45]    →0
[46]  T3:' THE INT IS OF TYPE 3 '
      ∇
```

LIST OF REFERENCES

1. Ogawara, M., "A Note on the test of serial correlation coefficients," Annals of Mathematical Statistics, v. 20, p. 115-118, 1950.

2. Hahn, G. J., "Prediction intervals for a normal distribution," G.E. Company Technical Information Series, Report No. 71-C-038, Dec 1970.

3. Choi, C. H., Prediction Intervals for Correlated Samples, M.S. Thesis, Naval Postgraduate School, Monterey, 1974.

4. Degroot, H. M., Probability and Statistics, p. 527-533, Addison Wesley, 1975.

5. Hogg and Craig, Introduction to Mathematical Statistics, MacMillan, 1965.

6. Hickman, J. C., "Preliminary Regional Forecasts for the outcome of an estimation problem," Journal of American Statistical Association, v. 58, p. 1104-1112, 1963.

7. Krishnaih, P. R. and Murthy, V. K., "Simultaneous tests for trend and serial correlations for Gaussian Markov residuals," Econometrica, v. 34, p. 472-480, 1966

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4.	Professor Toke Jayachandran Code 53Jy Department of Mathematics Naval Postgraduate School Monterey, California 93940	1
5.	LCDR T S Murthy Directorate of Naval Education and Meteorology Sena Bhawan, New Delhi India 110011	4